PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12Q 1/68, G01N 33/53

A1

(11) International Publication Number: WO 90/13668

(43) International Publication Date: 15 November 1990 (15.11.90)

(21) International Application Number: PCT/US90/02485

(22) International Filing Date: 4 May 1990 (04.05.90)

(30) Priority data:
348,350 5 May 1989 (05.05.89) US

(71) Applicant: LIFECODES CORPORATION [US/US]; Saw Mill River Road, Valhalla, NY 10595 (US).

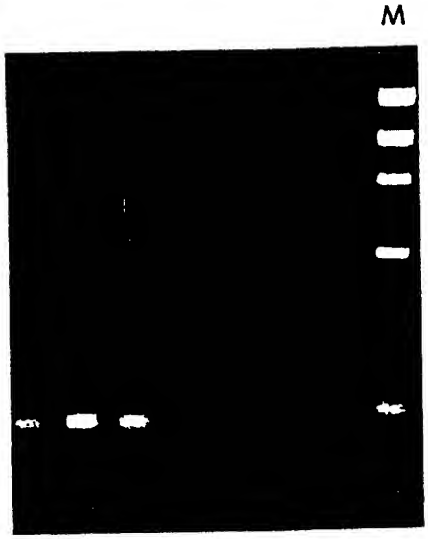
(72) Inventors: FISCHER, Stuart, G. ; 132 West 78th Street, New York, NY 10024 (US). IP, Nancy, Yuk-Yu ; 23 Emery Drive, Stamford, CT 06902 (US). SHALER, Robert, C. ; 30 Spring Street, Flemington, NJ 08822 (US). VAN DE STADT, Ingrid, L. ; 43 Prospect Avenue, Ossining, NY 10562 (US).

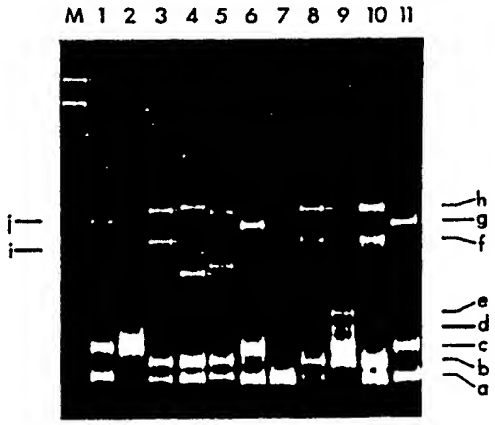
(74) Agent: MISROCK, S., Leslie; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: METHOD FOR GENETIC ANALYSIS OF A NUCLEIC ACID SAMPLE





(57) Abstract

The present invention provides a method for the genetic analysis of a nuclei acid sample comprising: (a) forming homoduplexes and heteroduplexes from at least one polymorphic region in said nucleic acid sample wherein the copy number of all of the variants of said polymorphic region in said nucleic acid sample has been amplified, and (b) differentiating said homoduplexes and said heteroduplexes by means of a nondenaturing gel. The method is very reliable and easy to perform, and can be utilized to analyze any type of polymorphism. The use of a labelled probe is not essential in order to differentiate the homoduplexes and heteroduplexes. Also, if the nucleic acid sample is contaminated with nucleic acid material from other individuals, the method permits such contamination to be readily apparent. Such method can be utilized to, *inter alia*, diagnose genetic disorders, establish identity and establish parentage.

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Monaco
AU	Australia	FI	Finland	MG	Madagascar
BB	Barbados	FR	France	ML	Mali
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Fasso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GR	Greece	NL	Netherlands
BJ	Benin	HU	Hungary	NO	Norway
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo			SN	Senegal
CH	Switzerland	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
DE	Germany, Federal Republic of	LK	Sri Lanka	TG	Togo
DK	Denmark	LU	Luxembourg	US	United States of America

METHOD FOR GENETIC ANALYSIS OF A NUCLEIC ACID SAMPLE

1. FIELD OF THE INVENTION

The present invention relates to a method for genetic analysis. Such method can be utilized to, inter alia, diagnose genetic disorders, establish identity and establish parentage.

2. BACKGROUND OF THE INVENTION

A polymorphism is the occurrence in the same population of two or more alleles at a genetic locus wherein the frequency of the most frequent allele does not exceed 99%. J.F. Gusella, DNA Polymorphism and Human Disease, Ann. Rev. Biochem., 55, 831-854 (1986). Polymorphisms are ultimately due to a difference in the primary sequence of genomic DNA in a population.

Polymorphisms can be utilized to differentiate between two copies of a particular locus in a genome. The ability to make such differentiation can be utilized to, inter alia, diagnose genetic disorders, e.g. sickle cell anemia and β-thalassemia, establish identity, e.g. forensic analysis, and establish parentage, e.g. paternity.

The use of polymorphisms to establish identity or establish paternity is based on the power to exclude, i.e. the statistical likelihood that two individuals will have the same allele for any given polymorphism. This likelihood is dependent upon not only the number of different alleles that exists but also on the frequency

with which each of the alleles occurs in the relevant population. Clearly, the greater the number of alleles, the greater the power to exclude.

Polymorphisms can be detected by a number of methods including direct sequence analysis and measurement of physical parameters of the nucleotide sequence. See L.S. Lerman et al., Sequence-Determined DNA Separations, Ann. Rev. Biophys. Bioeng., 13, 399-423 (1984).

One of the most straight forward and most frequently utilized methods utilizes restriction endonuclease digestion. The method is based upon the fact that restriction enzymes recognize specific nucleotide sequences and that changes in a nucleotide sequence, such as a base insertion or deletion, can result in the appearance or disappearance of a particular restriction endonuclease cleavage site, thereby altering the size of fragments generated from a given region. Differences in the size of fragments resulting from the digestion of the corresponding region of DNA from homologous chromosomes have been termed restriction fragment length polymorphisms or "RFLPs." See J.F. Gusella, DNA Polymorphism and Human Disease, Ann. Rev. Biochem., 55, 831-854 (1986).

The RFLPs are generally detected by restriction enzyme digestion of the nucleic acid sample, followed by gel fractionation of the resulting fragments, transfer of the fragments to a solid support, e.g., a filter, (Southern blotting), and hybridization to a labeled probe that is specific for the polymorphism in question, thereby providing a pattern that is characteristic of the nucleic acid sample. See D. Botstein et al., Construction of a Genetic Linkage Map in Man Using

Restriction Fragment Length Polymorphisms, Ann. J. Hum. Genet., 32, 314-331 (1980) and M.H. Skolnick, Strategies for detecting and characterizing restriction fragment length polymorphisms (RFLP's), Cytogenet. Cell Genet., 32, 58-67 (1982). As an alternative, one can utilize such labeled probe to directly hydridize to the fragments in the gel and, therefore, Southern blotting is not necessary. See M. Purrello et al., Anal. Biochem., 128, 393-397 (1983) and J. Gusella, Adv. Exp. Med. Biol., 154, 153-164 (1982).

The pattern resulting from such RFLP analysis can be compared to a known pattern in order to, inter alia, diagnose genetic disorders, establish identity or establish parentage.

A major limitation of such method is that only those polymorphisms that result in the appearance or disappearance of a restriction endonuclease cleavage site can be analyzed. Another limitation of the method is that a step utilizing a labeled probe in order to detect the polymorphism is required.

Another type of polymorphism wherein one can utilize a RFLP to detect the polymorphism is based not on the creation or destruction of a specific restriction endonuclease cleavage site, but on what has been referred to as a "hypervariable region." The hypervariable region consists of tandem repeats of a short nucleotide sequence (or "minisatellite") and polymorphism results from allelic differences in the number of tandem repeats. The resulting minisatellite length variation can be detected by using any restriction endonuclease that does not cleave the tandem repeat. See A. Jeffreys, Hypervariable

"minisatellite" regions in human DNA, Nature 314, 67-73 (1985) and A. Jeffreys et al., Individual-specific "fingerprints" of human DNA, Nature 316, 76-79 (1985).

Hypervariable regions have been detected in the myoglobin gene, the zeta-globin pseudogene, the insulin gene and the X-gene region of hepatitis B virus. See Y. Nakamura et al., Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping, Science, 235, 1616-1622 (1987) and G. Bell, The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences, Nature, 295, 31-35 (1982).

An allele-specific oligonucleotide (ASO) probe has been utilized to detect a polymorphism. The principle upon which an ASO probe works is referred to as "differential hybridization," which is based on the ability of the ASO probe to hybridize, under the appropriate conditions, only to those sequences to which it is perfectly matched. Thus, a single base mismatch is sufficiently destabilizing so as to prevent hybridization. Such method has been utilized to diagnose sickle cell anemia, β-thalassemia, as well as for the detection of HLA DNA polymorphisms. See B.J. Conner et al., Detection of sickle cell β^S-globin allele by hybridization with synthetic oligonucleotides, Proc. Natl. Acad. Sci. USA, 80, 278-282 (1983), V.J. Kidel et al., α-1 antitrypsin deficiency detection by direct analysis of the mutation in the gene, Nature, 304, 230-234 (1983) and R.K. Saiki et al., Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes, Nature, 324, 163-166 (1986).

More recently, such differential hybridization method was improved through the use of the polymerase chain reaction (PCR) procedure. The PCR procedure is a process for amplifying the copy number of a desired specific nucleic acid sequence contained in a nucleic acid or mixture thereof. The PCR procedure comprises treating separate complementary strands of the nucleic acid with a molar excess of two oligonucleotide primers, and extending the primers to form complementary primer extension products, which act as templates for synthesizing the desired nucleic acid sequence. The steps of the reaction can be repeated as often as is desired. See United States Patent 4,683,202, entitled, "Process for Amplifying Nucleic Acid Sequences."

The differential hybridization method utilized in combination with the PCR procedure is carried out by amplifying the copy number of that portion of the DNA to be analyzed, e.g. β-globin DNA, fixing such amplified DNA onto a filter, e.g. a nitrocellulose filter, contacting the fixed DNA with a labelled ASO probe under hybridization conditions followed by washing away the unhybridized ASO probe. The ASO probe is then detected by means of its label. Such type of assay is commonly referred to as a "dot blot analysis." See R.K. Saiki et al., Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes, Nature, 324, 163-166 (1986), J.A. Todd et al., HLA-DQβ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus, Nature, 329, 599-604 (1987) and T.L. Bugawan, The Use of Non-radioactive Oligonucleotide Probes to Analyze Enzymatically Amplified DNA for Prenatal Diagnosis and Forensic HLA Typing, BIO/TECHNOLOGY, 6, 943-947 (1988).

One advantage of the use of the PCR procedure with genetic analysis, in contrast to RFLP analysis, is that such procedure can be utilized with minute quantities of the nucleic acid sample.

Another advantage of the use of the PCR procedure with genetic analysis is that such procedure can be utilized with a degraded nucleic acid sample, so long as the portion of the nucleotide sequence to be amplified has not been degraded. In contrast, an RFLP analysis, which generally utilizes larger nucleotide fragments than are utilized with the PCR procedure, is less likely to be useful with a degraded nucleic acid sample. This is due to the fact that the degradation affects the size of such fragments and, therefore, the results of the genetic analysis.

Differential hybridization is an extremely unreliable method because the experimental conditions must be so stringent so as to permit one to differentiate a perfectly matched hybrid from a hybrid with a one base mismatch. Such stringent experimental conditions are extremely difficult to control, and when not maintained, result in non-specific hybridization, thereby producing a false positive result.

Similarly, a dot blot analysis is extremely unreliable. It is subject to a multitude of experimental conditions, any one of which could alter the result. Moreover, a separate dot blot analysis is required to analyze each allele of each polymorphism, thereby rendering the analysis very time consuming. Furthermore, in a dot blot assay it may be very difficult for one to compare nucleic acid samples, which is required when one attempts to establish identity, because each dot may contain a different number of copies of target DNA.

This affects the intensity of the signal and, therefore, one may not know whether or not a given intensity is due to the small quantity of target DNA on the dot or background.

Yet another drawback of a dot blot analysis, especially when utilized to establish identity, is that the nucleic acid sample may be contaminated with varying amounts of nucleic acid material from several individuals. In a dot blot analysis one cannot readily determine whether or not the dot contains nucleic acid material from more than one individual, thereby resulting in an erroneous interpretation of the data. For example, the dot blot analysis could result in several dots of high intensity and several dots of low intensity. One cannot readily determine whether or not the low intensity dots are a negative result from a first individual or a positive result from a small quantity of a second individual's nucleic acid material that is also present in the nucleic acid sample.

Accordingly, there is the need for a method for genetic analysis that can be utilized for all types of polymorphisms and that produces very reliable results. In addition, there is the need for a reliable method for genetic analysis of a nucleic acid sample wherein the nucleic acid sample is contaminated with a second individual's nucleic acid material.

3. SUMMARY OF THE INVENTION

The present invention provides a method for the genetic analysis of a nucleic acid sample comprising:

(a) forming homoduplexes and heteroduplexes from at least one polymorphic region in said nucleic acid sample wherein the copy number of all of the variants of said polymorphic region in said nucleic acid sample has been amplified, and

(b) differentiating said homoduplexes and said heteroduplexes by means of a nondenaturing gel.

The method is very reliable and easy to perform, and can be utilized to analyze any type of polymorphism. The use of a labelled probe is not essential in order to differentiate the homoduplexes and heteroduplexes. Also, if the nucleic acid sample is contaminated with nucleic acid material from other individuals, the method permits such contamination to be readily apparent.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A is a photograph of a stained agarose gel through which was run the product of the polymerase chain reaction described in Example 1.

Fig. 1B is a photograph of polyacrylamide gel slabs through which the PCR products were subjected to electrophoresis and stained with ethidium bromide as described in Example 1.

Fig. 2 is a photograph of the polyacrylamide gel slab obtained by the experiment reported in Example 2.

Fig. 3 is a photograph of an autoradiogram of dot blots previously hybridized with radioactive probes for DQα - 1, 3, 4 and 7.

Figs. 4A and 4B are photographs of stained polyacrylamide gel slabs through which PCR products from related individual's were subjected to electrophoresis.

Fig. 5 is a photograph of a stained polyacrylamide gel slab through which reamplified products of specific individual bands were subjected to electrophoresis.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method for the genetic analysis of a nucleic acid sample comprising:

(a) forming homoduplexes and heteroduplexes from at least one polymorphic region in said nucleic acid sample wherein the copy number of all of the variants of said polymorphic region in said nucleic acid sample has been amplified, and

(b) differentiating said homoduplexes and said heteroduplexes by means of a nondenaturing gel.

The principle upon which the present invention is based is most readily apparent and most useful, but the invention is not so limited, when the nucleic acid sample to be analyzed is derived from a sexually reproducing organism, e.g. a human. The genetic make-up of a sexually reproducing organism is such that it contains two sets of chromosomes, one set from each parent. Accordingly, for any given allele inherited from one parent there is a corresponding allele inherited from the other parent. Thus, for any given polymorphic region, two alleles exist. If each of such alleles is of the same nucleotide sequence, then that sexually reproducing organism is homozygous with respect to that polymorphic region and if each of such alleles is of a

different nucleotide sequence, then that sexually reproducing organism is heterozygous with respect to that polymorphic region.

The present invention amplifies the copy number of all of the variants - as defined hereinbelow - of at least one polymorphic region in the nucleic acid sample. The amplified nucleotide sequences are then permitted to anneal to each other. If heterozygosity exists, four resulting amplified double stranded fragments are formed:

1. a hybrid with no mismatches - a homoduplex -derived exclusively from the first allele;

2. a hybrid with no mismatches - a homoduplex - derived exclusively from the second allele;

3. a hybrid with mismatches - a heteroduplex - derived from the plus strand of the first allele and the minus strand of the second allele; and

4. a hybrid with mismatches - a heteroduplex - derived from the minus strand of the first allele and the plus strand of the second allele.

If the particular sexually reproducing organism is homozygous with respect to the particular polymorphic region that was amplified, then no heteroduplexes are formed.

The present invention utilizes the ability to differentiate between each of the homoduplexes and each of the heteroduplexes formed for a given nucleic acid sample. Such differentiation provides a unique pattern based upon all of the variants that are present in the nucleic acid sample for a given polymorphic region. The differentiation is by means of a nondenaturing gel whereby each different homoduplex and each different heteroduplex migrates differently, albeit it may be difficult to differentiate the bands formed by the two

homoduplexes. It should be noted that even if the two homoduplexes co-migrate, the present invention still can be utilized for genetic analysis because the two heteroduplexes migrate differently from the two homoduplexes. Thus, the relative position of each of the bands provides a unique pattern based upon all of the variants that are present in the nucleic acid sample for a given polymorphic region. Such pattern can then be compared to a known pattern to determine the genotype in order to, for example, diagnose a genetic disease, establish identity or establish parentage.

If the nucleic acid sample is homozygous with respect to the polymorphic region to be analyzed, then no heteroduplexes will be formed. However, homoduplexes will always be formed, which also can be utilized to form a pattern on a nondenaturing gel and then compared to a known pattern. Accordingly, the present invention is not limited to polymorphic regions that are heterozygous.

The present invention can be utilized for any genetic analysis of a nucleic acid sample. The term "genetic analysis" as used in the present invention includes any analysis of a nucleic acid sample that compares variations in a nucleotide sequence. The term "polymorphic region" as used in the present invention is a region of a nucleotide sequence that contains a variation of at least one nucleotide wherein, for the purpose of the present invention, each of such variations of the polymorphic region is a different "variant." An example of a variant for a eukaryotic organism is an allele. Thus, the present invention can be utilized not only to diagnose a genetic disease, establish identity and establish parentage but also to, for example, analyze various strains of a virus or a bacterium. Therefore,

the nucleic acid sample can be derived from, for example, a virus or a sexually reproducing organism, such as a mammal, including humans.

Nonlimiting examples of polymorphic regions that can be analyzed by the present invention include the polymorphic regions associated with the HLA loci, the apolipoprotein B gene, the human type II collagen gene , the Hras oncogene, the insulin gene, and the α-globin genes. See Bell et al., (1982) Nature (London) 295, 31-35, Proudfoot et al., (1982) Cell 31, 553-363, Capon et al., (1983) Nature (London) 302, 33-37, Boerwinkle et al., (1989) Genetics 86, 212-216, Stoker et al., Nucleic Acids Research, (1985) 13, 4613-4622.

The amplification step of the present invention permits the method of the invention to be performed with only a minute amount of the nucleic acid sample. Also, the present invention can be performed with a nucleic acid sample that has been degraded, so long as the polymorphic region to be amplified has not been degraded. The present invention also permits one to readily observe whether or not the nucleic acid sample is contaminated with nucleic acid material from a second individual.

5.1. Isolation of the Nucleic Acid Sample

Prior to carrying out the method of the present invention, the nucleic acid material is isolated from cells. The nucleic acid material can be obtained from, for example, blood, semen, tissue and amniotic fluid. The procedure can be carried out by conventional techniques. For example, the cells can be lysed with a lysing agent, e.g. a detergent such as sodium dodecyl sulfate. This results in the nucleic acid material being

accessible so that it can be purified from the cellular debris. Purification is carried out by standard techniques, for example, phenol extraction followed by alcohol precipitation.

5.2. Amplification of the Copy Number of All of the Variants of a Polymorphic Region

Once the nucleic acid sample has been isolated from the cell, the copy number of all of the variants of at least one polymorphic region in the nucleic acid sample is amplified. Of course, the copy number of more than one polymorphic region can be amplified, which, when utilized for the establishment of identity, results in a higher power of exclusion, as discussed hereinabove.

The copy number of all the variants of the polymorphic region can be amplified by any amplification technique known or to be developed in the future. One technique is to clone each of variants of the polymorphic region that are present in the nucleic acid sample. However, this technique is extremely tedious.

A preferred technique is the polymerase chain reaction (PCR) procedure. This technique is not only very straight forward to carry out but also results in all of the amplified copies of any given variant being the same length. This renders the differentiation between the resulting homoduplexes and heteroduplexes much easier. The PCR procedure is described in United States Patent 4,683,202, entitled, "Process for Amplifying Nucleic Acid Sequences." Yet another technique that can be utilized is a transcription-based amplification system (TAS). This technique produces multiple copies of RNA. See D.Y. Kwoh et al.,

Transcription based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format, Proc. Natl. Acad. Sci., 86, 1173-1177 (1989).

The length of the nucleotide sequence to be amplified depends on many factors. It is preferred that at least the entire polymorphic region be amplified. If less than the entire polymorphic region is amplified, then each amplified variant may not contain the variation in nucleotide sequence that is characteristic of each variant. Thus, when each of the variants is permitted to anneal, only homoduplexes will be formed, despite the fact that different variants are present, thereby not permitting each of the variants to be differentiated. However, in some circumstances it may not be necessary for each of the variants to be differentiated. For example, to diagnose a genetic disease, it is only essential to be able to differentiate the allele that is responsible for the disease from all of the other alleles rather than differentiating all of the alleles from each other. Thus, less than the entire polymorphic region may be able to be amplified and yet still permit one to differentiate the allele that is responsible for the disease from all of the other alleles, but not all of the alleles from each other. It is only essential that the length of the polymorphic region to be amplified be able to differentiate at least two variants.

To insure heteroduplex formation from all of the different variants, it is preferred that the polymorphic region to be amplified comprise the entire polymorphic region, and more preferably the entire polymorphic region and the nucleotide sequences flanking such polymorphic region.

If the flanking nucleotide sequences that are amplified are too long, the resultant amplified nucleotide sequence may be too long as compared to length of nucleotide sequence that is mismatched in the heteroduplex. This may render it very difficult to differentiate between the homoduplexes and heteroduplexes. On the other hand, if the length of the flanking nucleotide sequences to be amplified is too short, the complementary portions of the heteroduplex may not be long enough to form a stable heteroduplex.

If the PCR procedure is utilized for amplification, it is preferred that the primers be derived from the nucleotide sequences flanking the polymorphic region. This assures that one set of primers can be utilized to differentiate all of the variants of a given polymorphic region.

5.3. The Formation of Homoduplexes and Heteroduplexes

If after the amplification procedure the nucleotide sequences are in double stranded form, then the homoduplexes and heteroduplexes have already been formed. However, if after the amplification procedure the nucleotide sequences that were amplified are not in double stranded form, then the nucleic acid sample should be annealed by, for example, incubating the nucleic acid sample at room temperature or any other conventional technique. Also, since the amplified nucleotide sequences will more likely form homoduplexes - due to their enhanced stability - rather than heteroduplexes, it may be useful to melt the nucleotide sequences and let them reanneal under conditions that increase the likelihood of heteroduplex formation, e.g. lower

stringency. Also, if the PCR procedure is utilized, then after the last cycle, the nucleotide sequences should be melted and reannealed in order to permit maximum heteroduplex formation.

5.4. Differentiation of Homoduplexes and Heteroduplexes

Once each of the homoduplexes and each of the heteroduplexes have been formed, they then are differentiated by means of a nondenaturing gel. Thus, the nucleic acid sample is eletrophoresed on the nondenaturing gel and then each of the homoduplexes and each of the heteroduplexes are detected. The differentiation by means of the nondenaturing gel provides a unique pattern based upon all of the variants that are present in the nucleic acid sample, which have been amplified, for a given polymorphic region. Such pattern permits one to obtain information, e.g. diagnose a genetic disorder or establish identity.

A nondenaturing gel is utilized to differentiate between each of the homoduplexes and each of the heteroduplexes because it has been observed that each of the homoduplexes and each of the heteroduplexes migrates differently, with the mobility of a heteroduplex being retarded as compared to that of a homoduplex. Furthermore, it has been observed that such retardation is reproducible in that the relative position of each of the homoduplexes and each of the heteroduplexes remains constant over time, thereby more readily permitting the comparison of the pattern formed by one nondenaturing gel to the pattern formed by another nondenaturing gel.

Any type of nondenaturing gel can be utilized so long as the homoduplexes and heteroduplexes remain in duplex form and the homoduplexes and heteroduplexes can be differentiated, with a polyacrylamide nondenaturing gel being preferred.

The percentage of polyacrylamide and size of the gel should be sufficient to permit differentiation between the homoduplexes and heteroduplexes and is dependent upon the size of the homoduplexes and heteroduplexes and, therefore, must be empirically determined. For example, polyacrylamide gels of from about 6% to about 20%, preferably about 15%, were found to be sufficient to differentiate homoduplexes and heteroduplexes of about 250 base pairs.

After the gel has been run, the pattern formed by the homoduplexes and heteroduplexes can be detected by any technique, for example, staining the gel with ethidium bromide, thereby permitting visualization of the pattern. Another technique that can be utilized to detect the pattern is to utilize fluorescently labelled primers that can be detected in the gel during electrophoresis by means of a fluorescent detector. Thus, the pattern is recorded by the fluorescent detector.

It should be noted that another major benefit of the present invention is that because the copy number of the homoduplexes and heteroduplexes has been amplified, the mere staining of the gel permits one to readily differentiate such homoduplex band(s) and heteroduplex band(s) from the nonamplified nucleic acid material in the nucleic acid sample. Thus, there is no need for the use of a labelled probe, e.g., a radiolabelled or enzymatically labelled probe. However,

if desired, a labelled probe can be utilized after the amplification step by permitting the labelled probe to hybridize to the amplified nucleotide sequences.

In a less preferred embodiment, after the nondenaturing gel has been run, a denaturant can be added to melt the homoduplexes and heteroduplexes. One can then hybridize with a labelled probe under conditions to permit the labelled probe to hybridize to the homoduplexes and heteroduplexes or blot the homoduplexes and heteroduplexes on a filter and then hybridize with the labelled probe. The labelled probe is then detected.

Once the homoduplexes and heteroduplexes have been detected, the resulting pattern can be compared to a known pattern on the same gel or from another nondenaturing gel in order to obtain information, e.g. diagnose a genetic disease, establish identity or establish parentage. Also, the resulting pattern can be stored and utilized at some future date. If the resulting pattern is to be compared to a pattern derived from another nondenaturing gel, it may be desirable to utilize markers on the gel in order to normalize the results.

mM each; total volume 16 ul), primers (20 pmol/ul; 5 ul each) and Taq polymerase (5 U/ul; 0.5 ul). The Taq polymerase was obtained from Perkin Elmer Cetus, Emeryville, CA. PCR buffer (10X) was prepared to contain 500 mM KCl, 100 mM Tris-Cl at pH 8.3, 15 mM $MgCl_2$ and 0.1% (w/v) gelatin. The primers used to amplify the HLA-DQα locus in the PCR reaction were: 5′ - GTG - CTG - CAG - GTG - TAA - ACT - TGT - ACC - AG - 3′ and 5′ - CAC - GGA - TCC - GGT - AGC -AGC- GGT - AGA - GTT - G - 3′. The size of the amplified fragment using this set of primers is 242 base parts. Total volume of the above for the PCR reaction was 100 ul. To prevent evaporation, 50 ul of mineral oil was added on top of the sample.

PCR amplification was usually permitted to continue through about 25 cycles to produce a 10^6-fold amplification. In more detail, the procedure used was as follows. Amplification was allowed to proceed on a programmable heat block as follows: denaturation for 1 minute at 94°C; annealing for 1 minute at 55°C; and extension for 1 minute at 72°C, in total for 25 cycles.

The PCR products (10 ul) were run on a 3% agarose gel at 100 volts for 1 hour, using bacteriophage Phi X 174 DNA digested with Hae III as a marker through the gel. The stained gel is shown in the photograph reproduced in Fig. 1A. Lane M represents Phi X 174 phage DNA digested with Hae III. The other lanes represent the PCR products from the HLA-DQα locus of different individuals. They ran as a single 242 base pair fragment on the agarose gel.

Electrophoresis of the HLA-DQα locus DNA

A stock solution of 30% acrylamide was prepared from acrylamide (150 g) and bis-acrylamide (4 g) and brought up to volume (500 ml) with distilled water. To make 15% acrylamide (50 ml) for the electrophoresis medium, 30% stock acrylamide (25 ml) and 10X TAE (5 ml; 1X TAE = 0.04 M Tris at pH 7.4, 0.02M sodium acetate and 1 mM EDTA) were combined in distilled water (20ml). The pH was adjusted to 8.2, when necessary, with boric acid. Then ammonium persulfate (250 ul; 20%) and TEMED (53 ul; 100%) were added.

The acrylamide solution was poured between glass plates of a vertical slab gel electrophoresis unit with spacers (1.2 mm). A comb (1.2 mm; 16 teeth) was then inserted. The approximate size of the resulting gel is 16 x 17 cm. When ready to be used, the slab gel unit was placed in a TAE - filled aquarium tank having controls for temperature adjustment, wherein the temperature was adjusted to 25°C.

Depending on the quantity of the DNA in the sample, an aliquot (10-30ul) of the PCR product was used for the analysis. To increase heteroduplex formation, some PCR products were heat denatured at 94°C for 5 minutes and then allowed to reanneal at room temperature for 5 minutes. Following the addition of 5X running dye (5-6 ul; 2.5% Ficoll 400, 0.25% bromophenol blue, 0.25% xylene cyanol, 50 mM EDTA and 5X TAN) the resulting samples were loaded into the wells of the acrylamide gel so-prepared, the gels were electrophoresed for 15 hours at 150 volts with circulating buffer (25°C). The next day, the DNA in the gel was visualized by

staining with ethidium bromide (0.5 - 1.0 mg/l). The gels were then examined and photographed under ultraviolet illumination.

When the PCR products were run on the polyacrylamide gel (15%), they show distinct patterns for different individuals. The resulting stained gel is shown in the photograph reproduced as Fig. 1B. Approximately 14 distinct patterns have been observed thus far. Lane M represents Phi X 174 phage DNA digested with Hae III. Lanes 1 - 11 represent PCR products of samples 1 - 11 as indicated in Table 1.

To investigate the possibility that the polymorphic pattern seen on Fig. 1B might have been due to some non-specific fragments of varying sizes that were formed during the PCR reaction, the following experiment was performed.

Instead of running the PCR products on a polyacrylamide gel, some of the 242 base pair fragment from Fig. 1A was excised from the agarose gel as a thin agarose slice, placed horizontally across a 15% polyacrylamide gel and then subjected to electrophoresis. The result obtained is shown in Fig. 2.

It can be seen that the polymorphic pattern persists when the starting material is a single 242 base pair fragment. For example, compare lane 11 of Fig. 2 with lane 11 of Fig. 1B.

Example 2

Genotype Assignment

In order to determine which HLA-DQα alleles are represented by the bands seen on the polyacrylamide gel, the gel results were compared with

the typing done by the dot blot procedure. In some instances, genotype assignment was confirmed by sequence analysis.

Acrylamide Gel Electrophoresis

Electrophoresis was performed here as described in Example 1.

Dot Blot Procedure

MSI - nylon filters (Micron Separations, Inc., Westborough, MA) (cut to 9 cm x 13 cm) were rinsed in distilled water, soaked in 6X SSC for 10 minutes and the wet filter was then placed on a dot blot apparatus (Bethesda Research Laboratories, Inc., Bethesda, MD) which was attached to a vacuum pump. An aliquot (20 ng; approximately 1 ul) of the PCR product was brought up to 375 ul with distilled water. To this was added 5 M NaCl (100 ul) and 10 M NaOH (25 ul) to form a reaction mixture. The reaction mixture so-formed was then loaded into the slots of the dot blot apparatus with the vacuum pressure set at about 150 mm Hg. The sample is drawn by vacuum through the filter within about a minute. Once the slots of the apparatus are empty, neutralization solution (500 ul; 0.5 M Tris; pH 7.4; 2.5M NaCl) was introduced into each slot. Once the slots were empty, 500 ul of 2X SSC was introduced into each slot. After the slots were empty, the vacuum was shut off, the dot blot apparatus was disassembled and the filter was rinsed in 2X SSC. The filter was then air dried for about 15 minutes, baked in a vacuum oven (90°C; 2 hours) and subjected to ultraviolet irradiation (90 minutes). Each

dot blot was produced in quadruplicate, in order to hybridize each dot blot to a different probe (HLA-DQα_ 1, 3, 4 and 7).

The following oligomer probes were treated as described:

DQA 1 = TTC - AGC - AAA - TTT - GGA - GGT - T

DQA 3 = TGT - TTG - CCT - GTT - CTC - AGA - C

DQA 4 = TTC - CGC - AGA - TTT - AGA - AGA - T

DQA 7 = TTC - CAC - AGA - CTT - AGA - TTT - G

For the initial description and further information about probes DQA 3, 4 and 7, see Saiki et al., Nature 324:163-166 (1986).

An aliquot containing 1 μg of each of these four oligomers was combined with distilled water (29 ul), 4 ul of a mixture of spermidine (10 mM), Tris-Cl (0.2 M; pH 9.5), and EDTA (1 mM) to a total volume of 34 ul. This mixture was heated to 72°C for 2 minutes and quick chilled on ice.

To this was added 5 ul of 10x blunt-end kinase buffer (0.5 M Tris-Cl at pH 9.5, 0.1 M $MgCl_2$, 50 mM DTT and 50% glycerol). Then, T4 polynucleotide kinase (2 ul; 10,000 U/ml), ^{32}P gamma ATP (10 ul; 10 uCi/ul; Amersham; 5000 Ci/mmol) were added and the mixture was incubated for 30 minutes at 37'C.

To this was added EDTA (0.5 M; 2 ul) and the resultant was extracted with an equal volume (53 ul) of phenol/chloroform mixture (1:1; v/v). The volume was brought up to 90 ul with distilled water. A spin column consisting of Sephadex G-25 (1 ml) was prepared by centrifugation in an Eppendorf tube at 3000 rpm for 2 minutes. The sample is added to the column, which is

then centrifuged again at 2000 rpm for 2 minutes. The labelled oligomer was then collected. The efficiency of labeling is usually about 50 - 60%.

Hybridization of Dot Blot

Samples treated as described above were prehybridized at 50°C in hybridization solution (15 ml; 5X SSPE, 10% PEG, 1 mg/ml heparin, 2% SDS and 250 ug/ml salmon sperm) for one hour in a shaking waterbath (1X SSPE is 180 mM NaCl, 10 mM NaH_2PO_4 and 1 mM EDTA at pH 7.4). The dot blots were then hybridized with about 20 x 10^6 CPM ^{32}P oligomer probe (20 uCi/ug specific activity) in fresh hybridization solution (10 ml) overnight at 50°C in a shaking waterbath. The background improved when the period of hybridization was shortened to 1 h. The following day, the blots were washed for 10 minutes at 37°C in 0.1 x SSPE and 0.1% SDS. Radioactive dots were visualized by autoradiography at -70°C, using X - Omat AR (Kodak) X - ray film and Cronex Lighting Plus intensifying screen (DuPont). Average exposure time was about 90 minutes.

Results

Dot - blot procedure results were compared with results obtained by the electrophoresis procedure of the present invention. For example, DNA samples from lanes 7-12 of Fig. 1B were analyzed by the dot blot method, as shown in Fig. 3. Sample 9 hybridized to probes HLA-DQα 1 and 4, whereas sample 10 hybridized to HLA-DQα 4 and 7, and sample 11 hybridized to HLA-DQα 1 and 7.

Thus, it was concluded that band a from samples 10 and 11 represents allele 7, band b from samples 9 and 10 represents allele 4, and band c from samples 9 and 11 represents allele 1. These bands a-c are all homoduplexes of the different alleles.

As discussed in Examples 3 and 4, the upper bands that migrated slower are heteroduplexes. For example, bands d and e are heteroduplexes formed by alleles 1 and 4, bands f and h are formed by alleles 4 and 7, and band g by alleles 1 and 7.

One could not distinguish between alleles 3 and 7, since they both run as band a. See Table 1 and Fig. 1B. However, they can be distinguished by their respective heteroduplexes. For example, compare bands i and j from sample 1 and band g from sample 11.

Table 1

Sample No.*	HLA-DQα typing
1	1, 3
2	1, 1′
3	4, 7
4	4, 3
5	4, 3′
6	1, 7
7	7, 7
8	4, 7
9	1, 4
10	4, 7
11	1, 7
18	3, 3
19	3, 7
20	1, 7
21	3, 3′
22	3′, 7
23	4, 3
24	4, 3′

* See Figs. 1B and 4B.

The assignment of alleles 3/7, 4 and 1 to bands a, b and c respectively were further confirmed by sequence analysis of DNA from the bands. The DNA sequences for the alleles are in agreement with the published sequences in Gyllenstein and Erlich, Proc. Nat. Acad. Sci., 85:7656 (1988). It should be noted that the alleles cited in this reference are numbered differently (i.e., HLA-DQα 1, 2, 3 and 4 in the reference correspond to HLA-DQα 1, 7, 4 and 3, respectively in this disclosure).

Aside from the 4 major classes of DQα alleles (1, 7, 4 and 3), there are subtypes within allele 1 and allele 3. These subtypes of allele 1 and allele 3 can be distinguished by the present invention. Examples of this are seen on Table 1 (denoted by 1 and 1', 3 and 3').

Example 3

Mendelian Inheritance of the Alleles

The Mendelian inheritance of the alleles, as represented by the bands on the gel, were examined.

For example, as shown in Fig. 4A, lanes 6, 7 and 8 represent the DQ α genotype-specific pattern obtained from the father, child and mother respectively. It can be seen that the pattern in the child has only one single homoduplex band, thus the child has inherited the same allele from the parents. In another family, the child (lane 10) has inherited a different allele from each of his parents (lane 9 and 10). To show that the pattern seen in the child resulted from homoduplex and heteroduplex formation, the following experiment was performed. Amplified DQα alleles from the father (lane 9) and from the mother (lane 11) were mixed together, allowed to heat denature and reanneal at room temperature

to form various recombinations of homoduplexes and heteroduplexes and then electrophoresed. As shown in the mixture lane (9 + 11), 3 homoduplex bands and 5 heteroduplex bands were obtained. The 4 bands seen in the child can be correlated with some of those seen in the mixture. Thus, in lane 10, the 2 lower bands are homoduplexes while the two upper bands are heteroduplexes formed from a specific combination of the parental homoduplexes depending on which allele is inherited.

Similar results were obtained with another family study, as shown in Fig. 4B. Lanes 18, 19, 20 and 24 represent the genotype - specific pattern obtained for the paternal grandfather, father, paternal grandmother and mother, respectively. When a similar mixing experiment was performed for the paternal grandparents, the specific pattern seen in the father can be correlated with some of the bands seen in the mixture lane (18 + 20). Similarly, the pattern obtained for the parental mixture (lane 19 + 24) shows that it contains all the homoduplex and heteroduplex bands seen in the children (lanes 21 - 23). Thus, the genotype-specific pattern seen in the children results from the formation of homoduplexes and heteroduplexes based upon the inheritance of specific alleles from the parents.

Example 4

Further proof that the pattern resulted from homoduplexes and heteroduplex formation is provided by the following experiment.

Individual bands representing homoduplexes or heteroduplexes were excised from polyacrylamide gel. DNA was eluted from the gel slice by incubation in 100 ul TE (10 mM Tris at pH 8.0 and 1 mM EDTA) on a rocking

platform for approximately 15 hours at 50°C. One microliter of the eluant was then subjected to PCR reaction in the same manner as described in Example 1 and subjected to electrophoresis. When band a from sample 1 (Fig. 1B) was reamplified, only band a was obtained (lane 1a in Fig. 5). When band c from sample 1 was reamplified, only band c was obtained (lane 1 c in Fig. 5). However, reamplification of either band i or band j from sample 1 resulted in four bands (a, c, i and j) as shown in lanes 1i and 1j of Fig. 5. These results show that bands a and c are homoduplexes while bands i and j are heteroduplexes. A similar result was obtained with homoduplex bands a and c and heteroduplex band g from sample 6, as shown in lanes 6 a, c and g. Similarly, reamplification of the heteroduplex band k from sample 21 (Fig. 4B) resulted in bands a and k. Sequence analysis shows that DNA from band k contains sequences from DQα alleles 3 and 3′. In addition, sequence analysis of the upper two bands from sample 23 (Fig. 4B) shows that they are the 2 heteroduplexes formed by allele 3 and allele 4.

What is claimed is:

1. A method for the genetic analysis of a nucleic acid sample comprising:

(a) forming homoduplexes and heteroduplexes from at least one polymorphic region in said nucleic acid sample wherein the copy number of all of the variants of said polymorphic region in said nucleic acid sample has been amplified, and

(b) differentiating said homoduplexes and said heteroduplexes by means of a nondenaturing gel.

2. The method of claim 1 wherein said amplification is carried out by the polymerase chain reaction procedure.

3. The method of claim 1 wherein said nondenaturing gel is a polyacrylamide gel.

4. The method of claim 1 which further comprises comparing said homoduplexes and said heteroduplexes that have been differentiated with another differentiated set of homoduplexes and heteroduplexes.

5. The method of claim 4 wherein said comparison is to diagnose a genetic disease.

6. The method of claim 4 wherein said comparison is to establish identity.

7. The method of claim 4 wherein said comparison is to establish paternity.

8. The method of claim 1 wherein the nucleic acid sample is derived from a sexually reproducing organism.

9. The method of claim 8 wherein said sexually reproducing organism is a human.

10. The method of claim 1 wherein said polymorphic region is a hypervariable region.

11. The method of claim 1 wherein the polymorphic region is selected from the group consisting of one or more base insertions, one or more base deletions, one or more base substitutions and combinations thereof.

12. The method of claim 1 wherein said polymorphic region is selected from the group consisting of the polymorphic regions associated with the human type II collagen gene, the apolipoprotein B gene, the insulin gene, the α-globin genes, the Hras oncogene and the HLA loci.

13. The method of claim 12 wherein the polymorphic region is derived from the HLA loci.

14. The method of claim 13 wherein the polymorphic region is derived from the HLA DQα locus.

15. The method of claim 1 wherein said polymorphic region is DNA.

16. The method of claim 1 wherein prior to said differentiation step said homoduplexes and said heteroduplexes are denatured and reannealed.

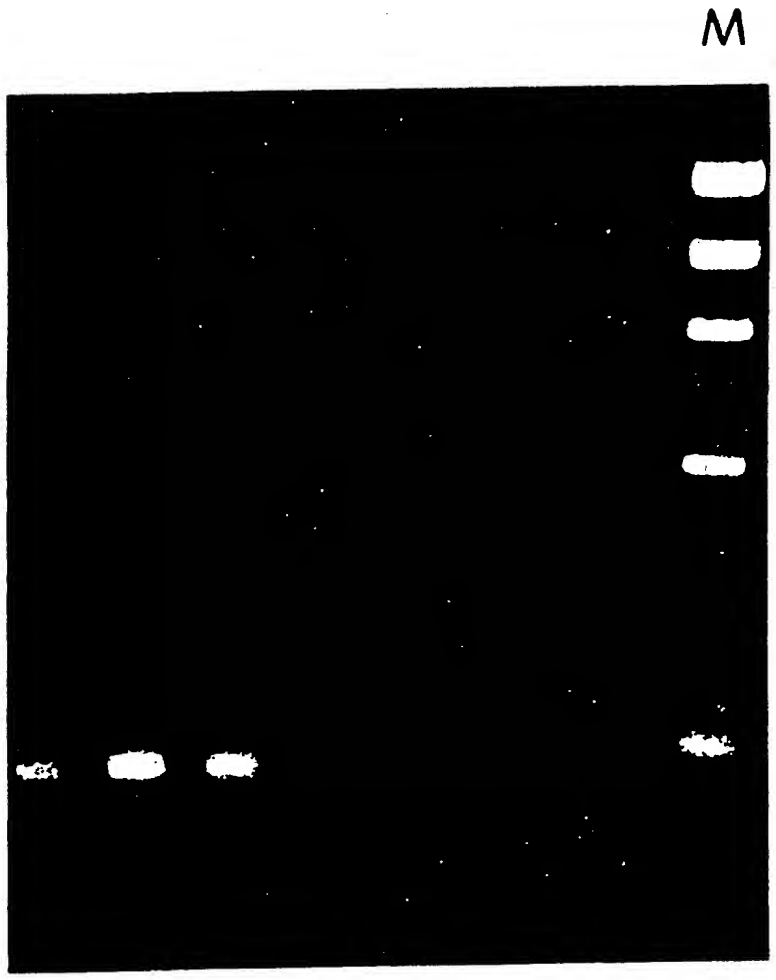


FIG. 1A

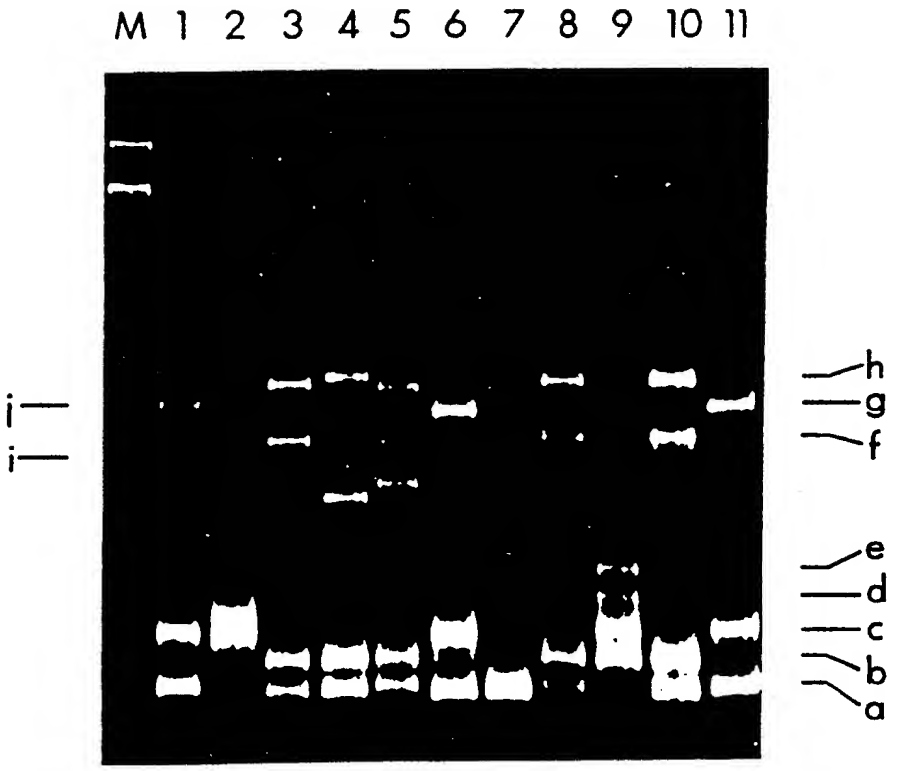


FIG. 1B

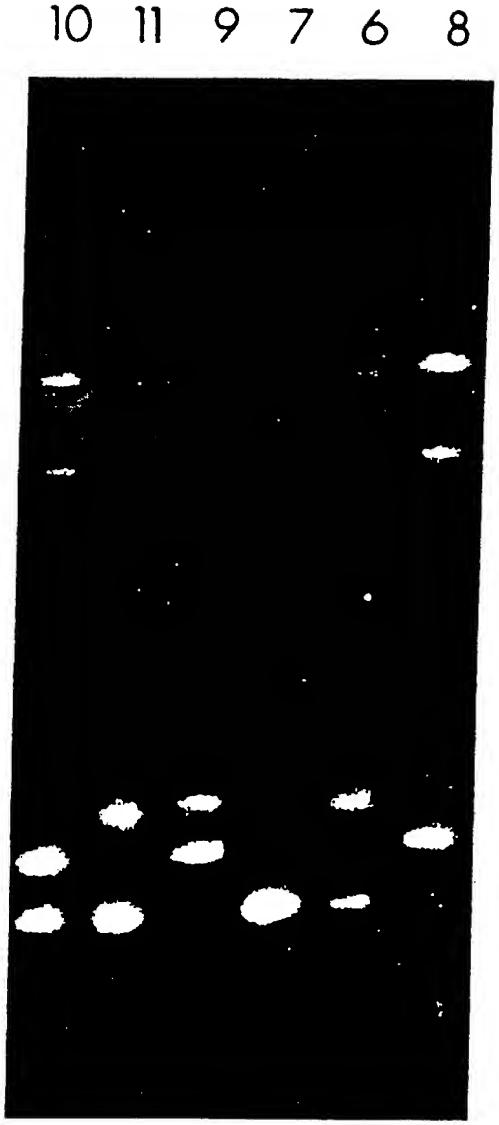


FIG. 2

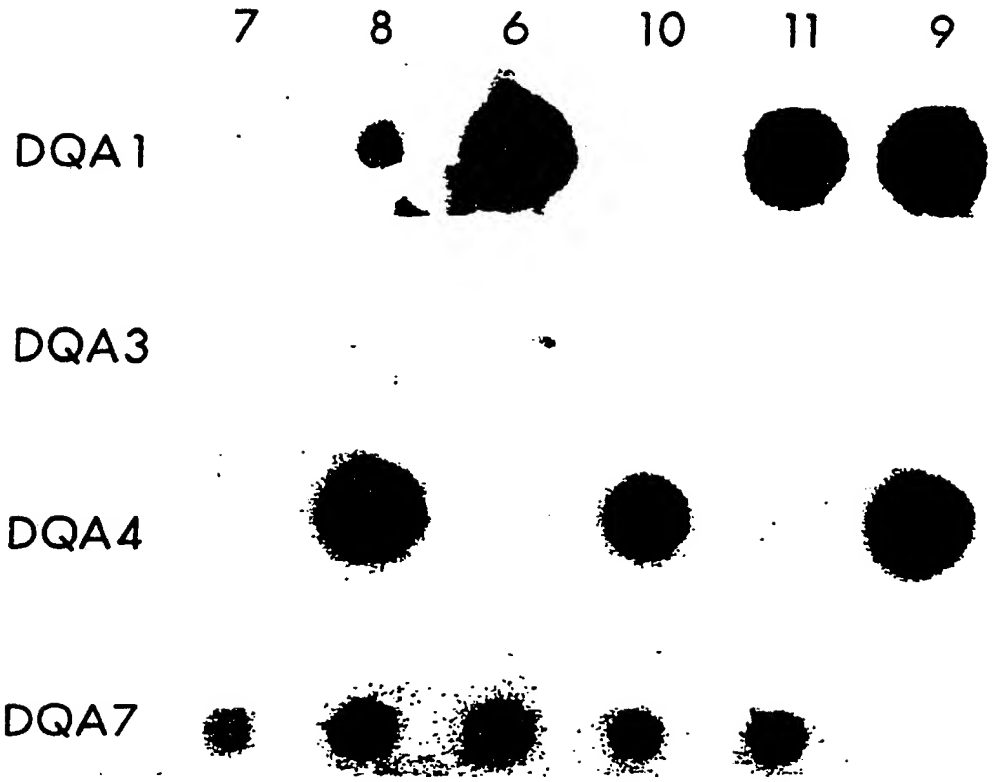


FIG. 3

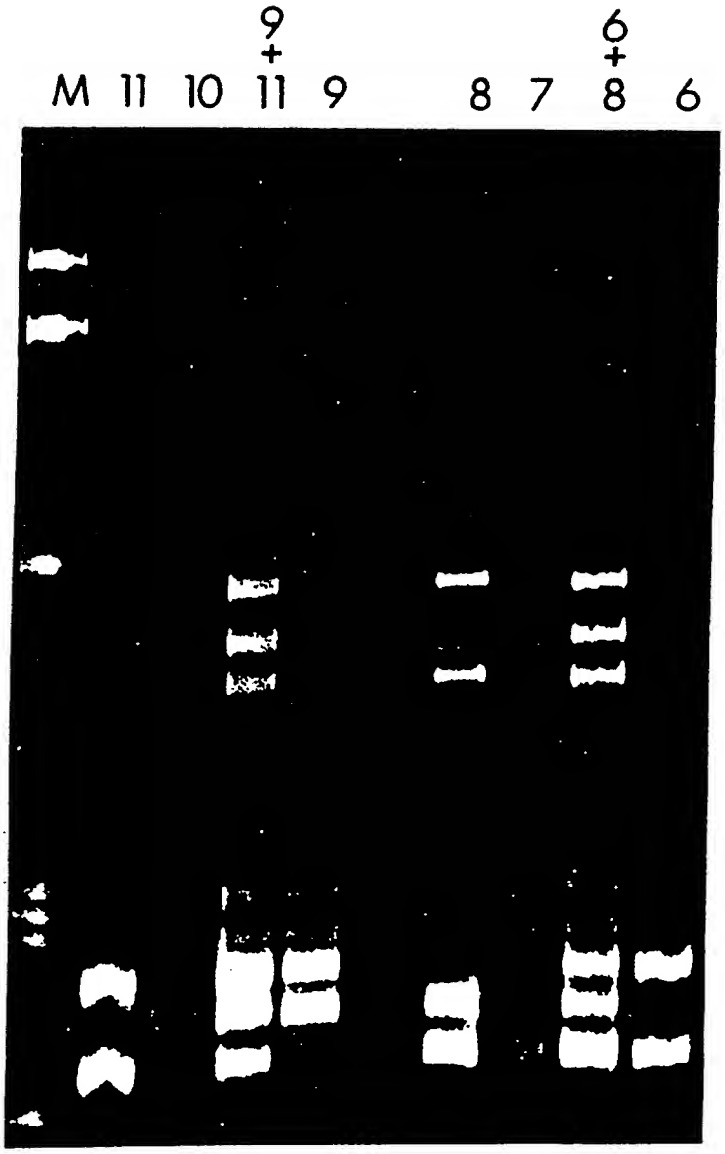


FIG. 4A

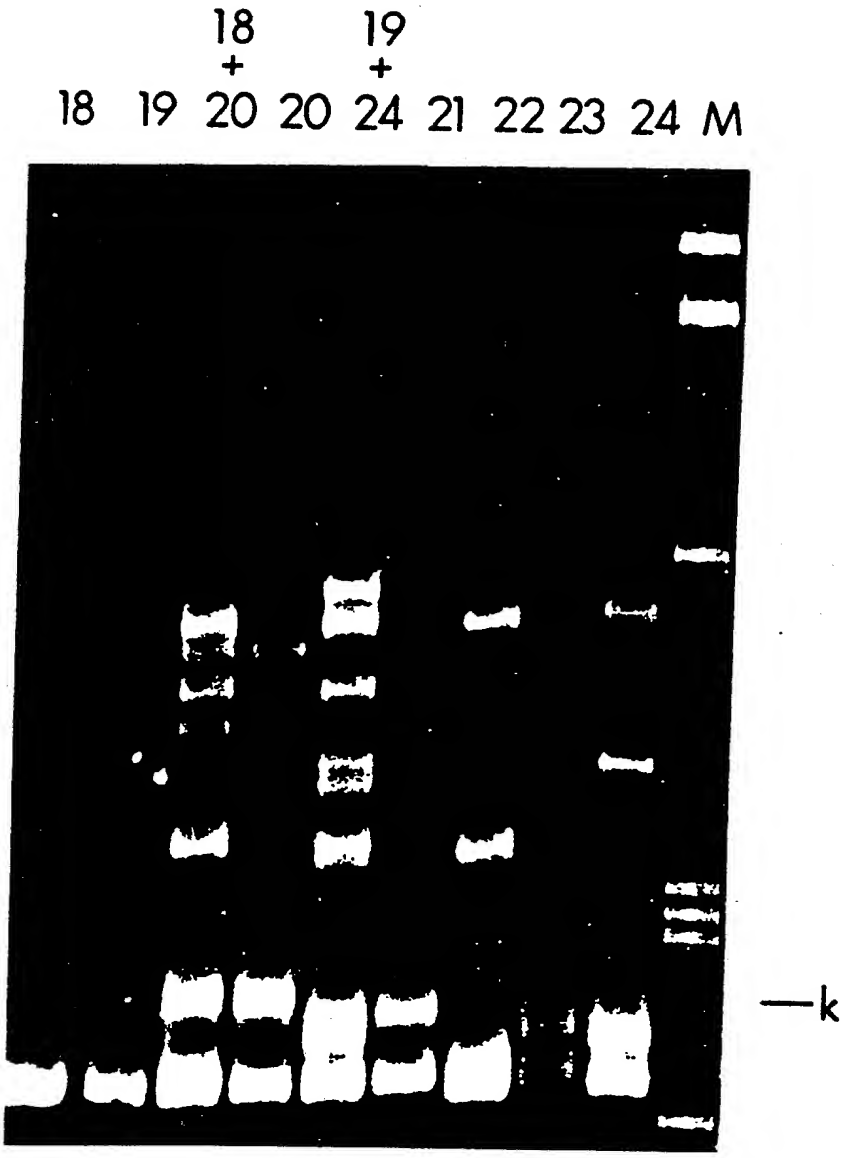


FIG. 4B

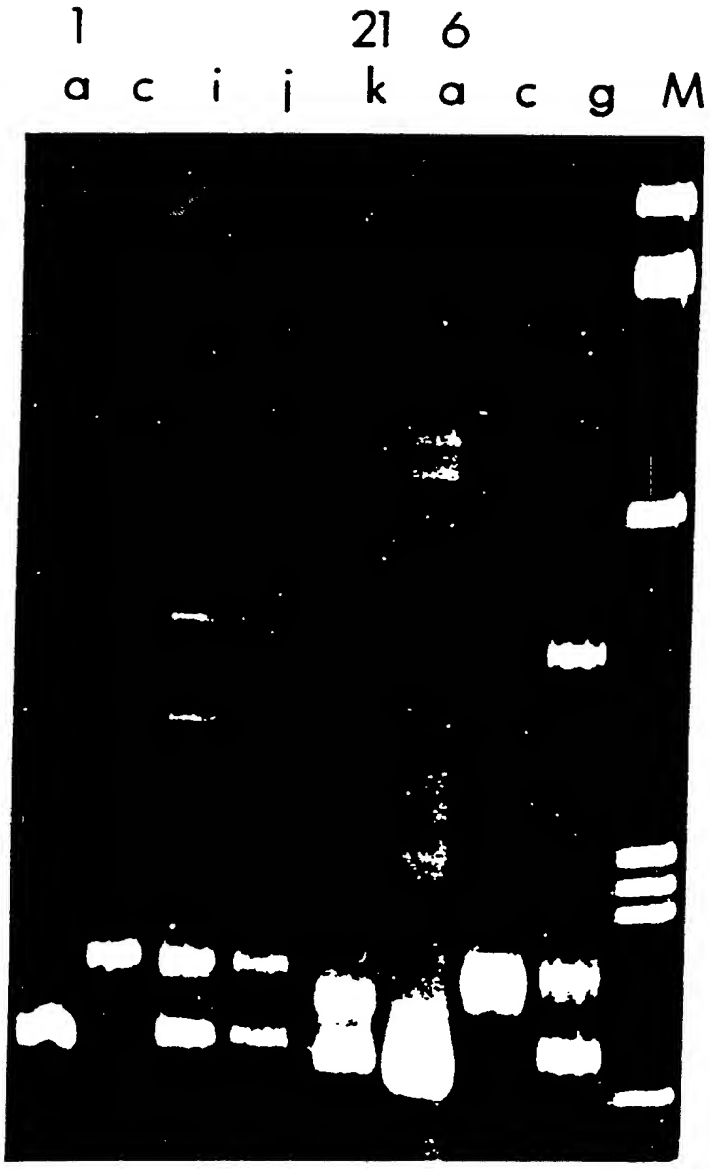


FIG. 5

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US90/02485

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 6

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C12Q 1/68; G01N 33/53

U.S.Cl.: 435/6 935/78

II. FIELDS SEARCHED

Minimum Documentation Searched 7	
Classification System	Classification Symbols
U.S. Cl.	435/6 935/78

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched 8

Cas Biosys & APS databases

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
Y	US, A, 4,683,195 (MULLIS et al.) 28 July 1987 see abstract.	1-16
Y	US, A, 4,794,075 (FORD et al.) 27 December 1988 see abstract.	1-16
Y	Nature, Volume 313 issued 07 February 1985 (London, England), Myers et al. "Detection of single base substitutions in total genomic DNA" see abstract.	1-16
Y	Proc. Natl. Acad. Sci. Volume 80 issued March 1983 (Washington, D.C. USA) Fischer et al. "DNA fragments differing by single base pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory" see abstract.	
Y	Science, Volume 229 issued 19 July 1985 (Washington, D. C. USA) Myers et al. "A general method for saturation mutagenesis of cloned DNA fragments" see pages 245-246.	1-16

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
18 July 1990	30 AUG 1990
International Searching Authority	**Signature of Authorized Officer**
ISA/US	For Scott A. Chambers Ngantho Nguyen

INTERNATIONAL DIVISION

Form PCT/ISA/210 (second sheet) (Rev.11-87)